



PROCEEDINGS

The 10th IMT-GT

**International Conference on Mathematics,
Statistics and its Application (ICMSA) 2014**

Knowing Nature Through Mathematical Science

October 14th - 16th, 2014
Kuala Terengganu



Predicting of Persistency and Permanency as well as Relations between Two Stages DNA Splicing Languages using C#

Mohammad Hassan Mudaber^a, Yuhani Yusof^a and Mohd Sham Mohamad^a

^a*Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang,
Lebuhraya Tun Razak, 26300 Gambang, Kuantan Pahang Darul Makmur*

Abstract. Splicing system is a system of studying the relation between mathematics and biology which was invented by Head in 1987. This connects three components: a set of four alphabets (nucleotides bases), a set of splicing rules (enzymatic operations) and a set of initial strings (double stranded-deoxyribonucleic acid or DNA). Persistent and permanent are two properties of splicing system which are more applicable in DNA process on generating the recombinant DNA strands at the second reaction. Therefore, a software is developed in presenting the persistency and permanency of recombinant DNA strands in order to optimize time and money. A user friendly interface called Microsoft Visual C Sharp (C#) is used in illustrating the behaviour of the resulted DNA molecules. Up to two stages DNA splicing system, this system works based on the developed mathematical theorems using Yusof-Goode (Y-G) approach. As an innovation in DNA molecular, this system can predict the persistency and permanency of two stages DNA splicing languages as well as relations between them.

Keywords: persistent, permanent, two stages Y-G splicing languages, C#

PACS: 87.14.gk, 87.14.ej

INTRODUCTION

Deoxyribonucleic acid (DNA) is a long sequence of repeating units of four nitrogenous bases namely adenine (*A*), guanine (*G*), cytosine (*C*) and thymine (*T*), which is used as a storage device for storing the genetic codes [1]. Naturally, DNA is the foundation of life, and it functions as proteins producer to form the genes which describe the organisms. According to structure, in all living organisms DNA exists as double-stranded sequences, where *A* is paired with *T* and *C* is paired with *G*. This pairing property between the bases is logical due to if two purine (double rings molecules) pairs to each other, the distance between two sugar-phosphate backbones will be increased. On the other hand, if two pyrimidines (single rings) pair to each other, then the distance between two backbones will be decreased. For this reason, the DNA molecules cannot make to keep its distance between two backbones same. However, a restriction enzyme can recognizes a particular sequence along the strand of DNA (restriction site) and then cut-off the sequence of DNA into fragments [2]. The restriction enzymes are categorized into two groups (produce sticky ends and blunt ends) according to their actions in order to cut the strand of DNA. Afterwards, the DNA fragments can only be combined to each other with complementary ends to generate new sequences of DNA.

Splicing system was defined in [3] as a formal device to generate the splicing languages by connecting the ideas from molecular biology and formal language theory. After splicing system had been introduced by Head, then it was defined and formulated in different notations by Paun, Pixton and Yusof-Goode (Y-G), see [4, 5, 6]. Head [3] and Gatterdam [7] focussed on the characteristics of splicing system, and introduced the persistency and permanency properties of splicing system. Recently, the persistent and permanent properties of splicing system as well as splicing languages that are produced by splicing system were extensively studied using Head, Paun and Y-G splicing system [8, 9, 10, 11]. In this research, the persistency and permanency of two stages DNA splicing languages, and the relations between stage one and stage two will be predicted by developing user friendly interface (UI) using Microsoft Visual C Sharp (C#) based on the provided mathematical theorems. In the next section, the preliminaries related to this study are discussed.

PRELIMINARIES

In this section, the fundamental definitions, which this research is based on them, are given. Persistent and permanent are two concepts that will be investigated in this paper. The results are presented and discussed in terms

of mathematical theorems and computer programming (C#). Persistent is a unique characteristics of splicing system which shows that the recombinant DNA molecules can be split by the existence of restriction enzymes when the manipulation process take place in stage two. Naturally, the restriction enzymes acting on double-stranded DNA can be determined in New England BioLab (NEB) according to their names, sequences, overhangs and types. However, there are no tools in NEB to predict whether the recombinant DNA strands will split if the restriction enzymes are chosen from actual biological sense or not. Therefore, to predict the persistency of two stages DNA splicing languages (recombinant DNA strands); the definition of persistent that was introduced by Head [3] is given below.

Definition 1: [3] Persistent

Let $S = (A, I, R)$ be a splicing system. Then S is persistent if for each pair of strings $ucxdv$ and $pexfq$, in A^* with (c, x, d) and (e, x, f) patterns of the same hands: if y is a sub segment of ucx (respectively xfq) that is crossing of a site in $ucxdv$ (respectively $pexfq$) then this same sub segment y of $ucxfq$ contains an occurrence of a crossing of a site in $ucxfq$. \square

Besides persistent, permanent is another characteristics of splicing system which plays pivotal role in fields of DNA recombination in order to predict whether the recombinant DNA strands will include the restriction enzymes sequences or not. Thus, this concept which its definition is almost similar to persistent is predicted using UI with respect to different combinations of DNA strands at the existence of restriction enzymes. The definition of permanent which was proposed by Gatterdam [4] is stated next.

Definition 2: [7] Permanent

Let $S = (A, I, R)$ be a splicing system. Then S is permanent if for each pair of strings $ucxdv$ and $pexfq$, in A^* with (c, x, d) and (e, x, f) patterns of the same hands: if y is a sub segment of ucx (respectively xfq) that is crossing of a site in $ucxdv$ (respectively $pexfq$) then this same sub segment y of $ucxfq$ is an occurrence of a crossing of a site in $ucxfq$. \square

In the next section, the results and discussions of this study are presented.

RESULTS AND DISCUSSIONS

This section discusses on the prediction of persistent and permanent characteristics of two stages splicing languages based on the results that are formulated mathematically as theorems by constructing a user friendly interface. Additionally, the relations between two stages DNA splicing languages are predicted based on the cutting sites of DNA strands. First of all, the user interface is introduced and then the procedure on how this software works is presented. Finally, the predictions of results using user friendly interface (UI) are discussed.

Introducing user interface (UI)

This section discusses on the structural appearance of UI. The executed form of UI involves six radio buttons which have been named 1 DNA, 2 DNA, 1 Enzyme, 2 Enzyme, 1 site and 2 sites, respectively. The first two radio buttons are used for choosing the number of DNA that user wants to insert, then the third and fourth radio buttons are used for choosing the number of restriction enzymes (RE), and the last two radio buttons are used to choose the number of cutting site(s) for the DNA strand. At the right side of UI, there are two text boxes for inserting DNA sequences and two text boxes for inserting RE sequences. However, if the user chooses only one DNA and one RE, then only one textbox for DNA and one textbox for RE will be appeared. The text box for inserting RE has three gaps, which are made based on the RE sequence. Therefore, from left to the right, the user should insert the left-handed context, crossing site and right-handed context to the first, second and third gaps, respectively. There is a reset, button which is used, if the user inserts wrong sequence of DNA or RE to re-insert the input. Finally, to check the results, there is a check button, for showing the persistency and permanency of two stages DNA splicing languages as well as relations between them.

Furthermore, to give more explanations to the user regarding DNA and its components, RE and cutting site of DNA, all necessity information is provided into the UI. If the user puts the cursor on the radio buttons belonging to 1DNA, 1RE and 1cutting sites of DNA on the interface, respectively then all related information will be appeared.

Procedure of Working UI

In this section, the procedure on how this software works is described. To use this UI, there are some steps to follow for predicting the persistency and permanency of two stages splicing languages as well as relations between stage one and stage two splicing languages. Biologically, this system predicts the paths of recombinant DNA molecules that might or might not be split by the existence of RE as well as to determine whether the new DNA fragments, will be produced at stage two or not. Since this software only can recognize and accept the four alphabets A, G, C and T , thus user has to insert the DNA sequences and RE that contain only these four letters. This software focuses on at most two initial strings (with one or two cutting sites) and two rules. Thus, the procedures of using this software are summarized below:

1. Choosing the number of DNA sequence (s)
2. Choosing the number of RE
3. Choosing the number of recognition site(s) for DNA strand(s)
4. Inserting the DNA sequence (s)
5. Inserting the RE sequence (s)
6. The system will check the number of cutting sites for DNA strand (s) as well as compatibility of the RE sequence (s) with DNA strand (s). However, if the user inserts the wrong sequence of DNA or inserts RE sequence that does not match with DNA strand, the system shows a message that insert DNA with one or two cutting sites, or RE does not match with DNA.
7. Check the checkbox to show the persistency and permanency or non-persistency and non-permanency of two stages splicing languages as well as relations between them.

Prediction of Results of UI

As Head (1987) defined, the languages that are generated by splicing system is called splicing languages. These splicing languages have some important properties such as persistent and permanent. Using Y-G splicing system these two characteristics of splicing languages up to two stages are investigated by providing mathematical theorems. Furthermore, the relations between stage one and stage two splicing languages are also studied. Therefore, the results that are obtained mathematically based on splicing rules properties and cutting sites of the initial DNA strands, are presented in below.

Theorem 1[12]: The set of two stages splicing languages which is produced by Y-G splicing system consisting two initial strings and two rules with disjoint crossing sites and palindromic sequences is persistent.□

Theorem 2 [12]: The set of two stages splicing languages which is produced by Y-G splicing system consisting two initial strings and two rules with disjoint non-palindromic crossing sites (or sequences) is persistent.□

Theorem 3 [12]: The set of two stages splicing languages which is produced by Y-G splicing system consisting two initial strings and two rules where the whole sequence of one rule is palindromic and the other is non-palindromic is persistent.□

Theorem 4: Let $S = (A, I, R)$ is a Y-G splicing system such that $I = \{\alpha axbcyd \beta\}$ and $R = \{(a, x, b : a, x, b), (c, y, d : c, y, d)\}$, $x \neq y$ and for x and y as palindromic strings, a complement with b and c complement with d and vice-versa and $a, b, c, d, x, y, \alpha, \beta \in A^*$. Then the two stages splicing languages are persistent.□

Theorem 5: The set of two stages splicing languages, that is produced by Y-G splicing system consisting two initial strings (with two cutting sites) and two rules with disjoint crossing sites and non-palindromic sequences, is persistent.□

Theorem 6: The set of two stages splicing languages, which is generated by Y-G splicing system consisting two initial strings (with two cutting sites) and two rules with disjoint crossing sites and palindromic sequences, is persistent.□

Theorem 7: The set of two stages splicing languages that is produced by Y-G splicing system consisting two initial strings (with two cutting sites) and two rules where the crossing of one rule is palindromic and the crossing of the other rule is non-palindromic, is persistent.□

Theorem 8 [13]: If each of the initial strings in a Y-G splicing system has only one recognition site, then no distinct splicing languages will be produced at stage two.□

Theorem 9 [13]: If each of the initial strings in a Y-G splicing system has two recognition sites, then the set of stage two splicing languages always contains the set of stage one splicing languages.□

According to the above results, a user friendly interface is developed using Microsoft Visual C Sharp (C#) to predict the persistency and permanency of two stages DNA splicing languages as well as the relations between stage one and stage two DNA splicing languages. Therefore, to predict the given theorems by UI easily, this section is divided into two subsections, namely persistency and permanency prediction and predicting the relations between two stages DNA splicing languages. Hence, the persistency and permanency prediction of splicing languages by developing UI is first discussed below.

Persistency and permanency prediction

As it is obvious, conducting laboratory experiments need to huge expenditure as well as time-consuming. Therefore, the persistent and permanent characteristics of two stages splicing languages are predicted by developing a user friendly interface. The purpose of constructing this UI is to avoid time and money consuming. Besides, to predict the persistency and permanency of splicing languages that will be produced by Y-G splicing system at stage one and stage two in terms of computer programming by using the provided mathematical theorems as stated above. It is noteworthy that if users want to predict the persistency and permanency of two stages splicing languages, they should follow the procedure correctly. To predict the persistency and permanency of two stages DNA splicing languages, the users can insert different possibilities of DNA strands and RE according to DNA strands and RE factors such as no of cutting site of DNA, palindromic and non-palindromic sequence of DNA, palindromic and non-palindromic sequences of RE, palindromic and non-palindromic crossing sites, identical and disjoint crossing sites, same left and right handed contexts, different left and right handed contexts, being left and right contexts complementary and non-complementary. Then, the software shows the result whether the two stages splicing languages are persistent and permanent or non-persistent and non-permanent as well as explains the reasons based on the proven theorems. In the following, the persistency and permanency of two stages DNA splicing languages are predicted using software at the existing of two DNA strands (with two cutting sites) and two restriction enzymes with disjoint palindromic sequences. The output is shown in **FIGURE 1**.

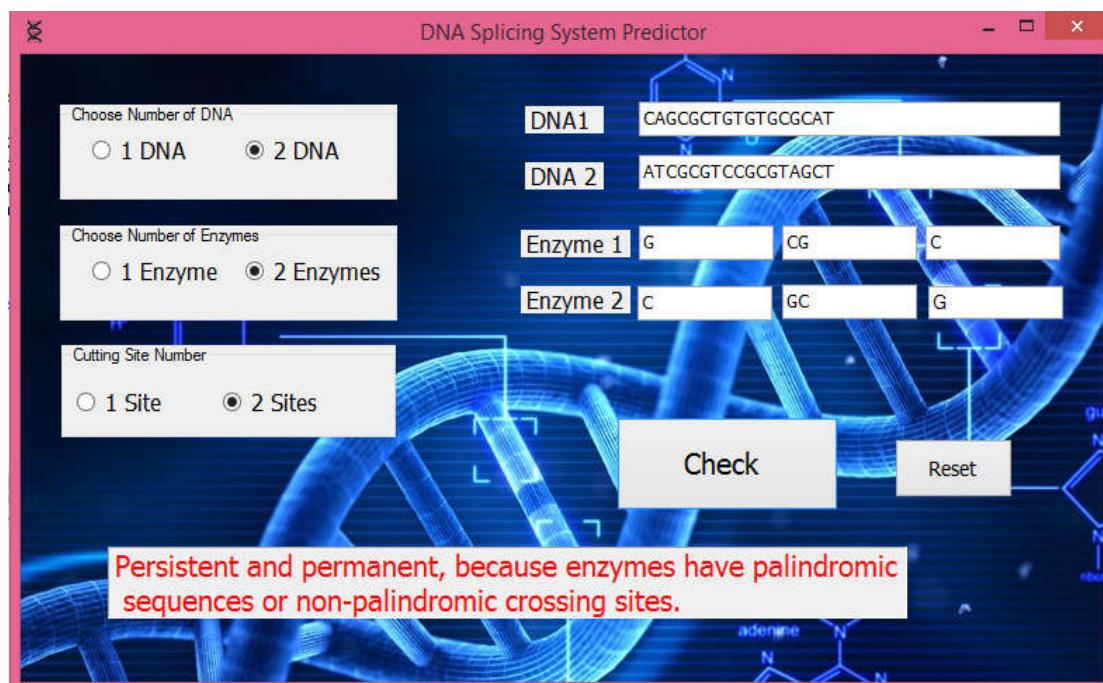


FIGURE 2. Persistency and permanency predictor via UI

However, the whole results that will be obtained by this software based on restriction enzymes properties and number of cutting sites of DNA strand are presented and summarized in the TABLE (1).

TABLE (1). Predicting the Persistency and Permanency of Two Stage Splicing Languages via UI

No. of RE	No. of DNA	RE Properties	DNA Properties	Results
1	1	Palindromic sequence	One cutting site	Persistent and permanent
1	1	Non-palindromic crossing site	One cutting site	Persistent and permanent
1	1	Palindromic crossing site with non-complementary left and right handed contexts	One cutting site	Non-persistent and non-permanent
1	2	Palindromic sequence	One cutting site	Persistent and permanent
1	2	Non-palindromic crossing site	One cutting site	Persistent and permanent
1	2	Palindromic crossing site with non-complementary left and right handed contexts	One cutting site	Non-persistent and non-permanent
2	2	Palindromic sequences and disjoint crossing sites	One cutting site	Persistent and permanent
2	2	Non-palindromic and disjoint crossing sites	One cutting site	Persistent and permanent
2	2	Palindromic disjoint crossing site with non-complementary left and right handed contexts	One cutting site	Non-persistent and non-permanent
2	2	Identical crossing site	One cutting site	Non-persistent and non-permanent

2	2	One of the RE has palindromic sequence and the other has non-palindromic (crossing site and contexts) sequence	Two cutting sites	Persistent and permanent
2	1	Palindromic sequence	Two cutting sites	Persistent and permanent
2	1	Non-palindromic crossing site	Two cutting sites	Persistent and permanent
2	1	Palindromic crossing site with non-complementary left and right handed contexts	Two cutting sites	Non-persistent and non-permanent
2	2	Palindromic sequences and disjoint crossing sites	Two cutting sites	Persistent and permanent
2	2	Non-palindromic disjoint crossing sites	Two cutting sites	Persistent and permanent
2	2	Palindromic disjoint crossing site with non-complementary left and right handed contexts	Two cutting sites	Non-persistent and non-permanent
2	2	Identical crossing site	Two cutting sites	Non-persistent and non-permanent
2	2	One of the RE has palindromic sequence and the other has non-palindromic (crossing site and contexts) sequence	Two cutting sites	Persistent and permanent

In the next section, the relations between two stages DNA splicing languages are predicted using UI.

Predicting the Relations between Two Stages DNA Splicing Languages

From mathematical consideration, it can be concluded that if the initial strings have only one cutting site then the set of stage one splicing languages is the same as stage two. However, if the initial strings have two cutting sites, then the set of stage one SL is a subset of stage two. In this section, based on this consideration the relations between two stages splicing languages are predicted by using UI. The constructed UI is able to predict the relations between stage one and stage two with respect to different combinations of initial strings under effect of appropriate rules. As a result, the relations between stage one and stage two DNA splicing languages are predicted using software at the existing of two DNA strands (with one cutting sites) and two restriction enzymes with disjoint palindromic sequences. The output is shown in **FIGURE 2**.

FIGURE 2. Predicting the Relation between Two Stages Splicing Languages via UI

However, the results that will be yielded from this software to predict the relations between stage one and stage two splicing languages based on the number of RE, DNA and cutting sites of DNA strands are summarized in table below.

TABLE (1). Predicting the Relations between Two Stages Splicing Languages via UI

No. of DNA	No. of RE	No. of DNA cutting site	Results
1	1	1	The sets of stage one and stage two Splicing Languages are the same.
2	1	1	The sets of stage one and stage two Splicing Languages are the same.
2	2	1	The sets of stage one and stage two Splicing Languages are the same.
1	1	2	The set of stage one Splicing Languages is a subset of Stage two Splicing Languages.
1	2	2	The set of stage one Splicing Languages is a subset of Stage two Splicing Languages.
2	1	2	The set of stage one Splicing Languages is a subset of Stage two SL.
2	2	2	The set of stage one Splicing Languages is a subset of Stage two Splicing Languages.

In the next section, conclusion of this paper is presented.

CONCLUSION

This paper discusses on programming aspect of DNA splicing languages. A user friendly interface is developed using Microsoft Visual C Sharp (C#) to predict the persistency and permanency of two stages DNA splicing languages as well as relations between them. This software works based on the provided mathematical theorems.

ACKNOWLEDGMENTS

The authors would like to express their heartfelt appreciation to Ministry of Education (MOE) of Malaysia and Research and Innovation Department of Universiti Malaysia Pahang for financial funding of this research article to the ICMSA2014 through Research Grant Vote: RDU 130354 and RAGS Vote: 131404. The first author also acknowledges Ministry of Higher Education (MOHE) of Afghanistan for his master scholarship and Kabul Education University (KEU) for study leave.

REFERENCES

1. S. Pandey, *Basics of Molecular Biology*, New Delhi: Anmol publications, 2009, pp. 85-89.
2. J. M. Walker, *Molecular Biology and Biotechnology*, RSC Publishing, 2009, pp. 1-3.
3. T. Head, *Bulletin of Mathematical Biology*, **49**, 737-759 (1987).
4. G. Paun, *Discrete Applied Mathematics*, **70**, 57-79 (1996).
5. D. Pixton, *Discrete Applied Mathematics* **69**, 101-124 (1996).
6. Y. Yusof, N. H. Sarmin, W. H. Fong, T. E. Goode and M. A. Ahmad, *An analysis of Four Variants of Splicing System*, AIP Conference Proceedings, American Institute of Physics, Melville, NY, 2013, pp. 888-895.
7. R. W. Gatterdam, *International Journal of Computer Math.* **31**, 63-67 (1989).
8. F. Karimi, N. H. Sarmin and W. H. Fong, *Australian Journal of Basic and Applied Sciences.* **5**, 20-24 (2011).
9. F. Karimi, N. H. Sarmin and W. H. Fong, *International Journal of Modern Physics.* **9**, 89-94 (2012).
10. F. Karimi, "Mathematical Modelling of Persistent Splicing System in DNA Computing ", Ph.D. Thesis, Universiti Teknologi Malaysia, 2013.
11. Y. Yuhani, N. H. Sarmin and W. H. Fong, "The Concept of Persistent and Permanent in Non Semi-Simple DNA Splicing System", AIP Conference Proceedings, American Institute of Physics, Melville, NY, 2014, 586-590.
12. M. H. Mudaber, Y. Yuhani and M. S. Mohamad, "Some Sufficient Conditions for Persistency and Permanency of Two Stages DNA Splicing Languages via Yusof-Goode Approach", AIP Conference Proceedings, American Institute of Physics, Melville, NY, 2014, 591-595.
13. M. H. Mudaber, Y. Yuhani and M. S. Mohamad, "Some Relations between Two Stages DNA Splicing Languages", AIP Conference Proceedings, American Institute of Physics, Melville, NY, 2014, 254-259.